Amendments to the Claims

This listing of claims replaces all prior versions and listings of claims in the application.

Listing of Claims

1. (Original) A method of synthesizing a set of digital speech samples corresponding to a selected voicing state from speech model parameters, the method comprising the steps of:

dividing the speech model parameters into frames, wherein a frame of speech model parameters includes pitch information, voicing information determining the voicing state in one or more frequency regions, and spectral information;

computing a first digital filter using a first frame of speech model parameters, wherein the frequency response of the first digital filter corresponds to the spectral information in frequency regions where the voicing state equals the selected voicing state;

computing a second digital filter using a second frame of speech model parameters, wherein the frequency response of the second digital filter corresponds to the spectral information in frequency regions where the voicing state equals the selected voicing state;

determining a set of pulse locations;

producing a set of first signal samples from the first digital filter and the pulse locations;

producing a set of second signal samples from the second digital filter and the pulse locations;

combining the first signal samples with the second signal samples to produce a set of digital speech samples corresponding to the selected voicing state.

2. (Original) The method of claim 1 wherein the frequency response of the first digital filter and the frequency response of the second digital filter are zero in frequency regions where the voicing state does not equal the selected voicing state.

3. (Original) The method of claim 2 wherein the spectral information includes a set of spectral magnitudes representing the speech spectrum at integer multiples of a fundamental frequency.

4. (Original) The method of claim 2 wherein the speech model parameters are generated by decoding a bit stream formed by a speech encoder.

5. (Original) The method of claim 2 wherein the voicing information determines which frequency regions are voiced and which frequency regions are unvoiced.

6. (Original) The method of claim 5 wherein the selected voicing state is the voiced voicing state and the pulse locations are computed such that the time between successive pulse locations is determined at least in part from the pitch information.

7. (Original) The method of claim 6 wherein the pulse locations are reinitialized if consecutive frames or subframes are predominately not voiced, and future determined pulse locations do not substantially depend on speech model parameters corresponding to frames or subframes prior to such reinitialization.

8. (Original) The method of claim 5 wherein the first digital filter is computed as the product of a periodic signal and a pitch-dependent window signal, and the period of the periodic signal is determined from the pitch information for the first frame.

9. (Original) The method of claim 8 wherein the spectrum of the pitch dependent window function is approximately equal to zero at all non-zero integer multiples of the pitch frequency associated with the first frame.

10. (Original) The method of claim 5 wherein the first digital filter is computed by:

determining FFT coefficients from the decoded model parameters for the first frame in frequency regions where the voicing state equals the selected voicing state;

processing the FFT coefficients with an inverse FFT to compute first time-scaled signal samples;

interpolating and resampling the first time-scaled signal samples to produce first time-corrected signal samples; and

multiplying the first time-corrected signal samples by a window function to produce the first digital filter.

11. (Original) The method of claim 10 wherein regenerated phase information is computed using the decoded model parameters for the first frame, and the regenerated phase information is used in determining the FFT coefficients for frequency regions where the voicing state equals the selected voicing state.

12. (Original) The method of claim 11 wherein the regenerated phase information is computed by applying a smoothing kernel to the logarithm of the spectral information for the first frame.

13. (Original) The method of claim 11 wherein further FFT coefficients are set to approximately zero in frequency regions where the voicing state does not equal the selected voicing state or in frequency regions outside the bandwidth represented by speech model parameters for the first frame.

14. (Original) The method of claim 10 wherein the window function depends on the decoded pitch information for the first frame.

15. (Original) The method of claim 14 wherein the spectrum of the window function is approximately equal to zero at all integer non-zero multiples of the pitch frequency associated with the first frame.

16. (Original)The method of claim 2 wherein the selected voicing state is a pulsed voicing state.

17. (Currently Amended) The method of ~~claims~~ claim 16 wherein the first digital filter is computed as the product of a periodic signal and a pitch-dependent window signal, and the period of the periodic signal is determined from the pitch information for the first frame.

18. (Original) The method of claim 17 wherein the spectrum of the pitch dependent window function is approximately equal to zero at all non-zero integer multiples of the pitch frequency associated with the first frame.

19. (Currently Amended) The method of ~~claims~~ claim 16 wherein the first digital filter is computed by:

determining FFT coefficients from the decoded model parameters for the first frame in frequency regions where the voicing state equals the selected voicing state;

processing the FFT coefficients with an inverse FFT to compute first time-scaled signal samples;

interpolating and resampling the first time-scaled signal samples to produce first time-corrected signal samples; and

multiplying the first time-corrected signal samples by a window function to produce the first digital filter.

20. (Original) The method of claim 19 wherein regenerated phase information is computed using the decoded model parameters for the first frame, and the regenerated phase

information is used in determining the FFT coefficients for frequency regions where the voicing state equals the selected voicing state.

21. (Original) The method of claim 20 wherein the regenerated phase information is computed by applying a smoothing kernel to the logarithm of the spectral information for the first frame.

22. (Original) The method of claim 20 wherein further FFT coefficients are set to approximately zero in frequency regions where the voicing state does not equal the selected voicing state or in frequency regions outside the bandwidth represented by speech model parameters for the first frame.

23. (Original) The method of claim 19 wherein the window function depends on the decoded pitch information for the first frame.

24. (Original) The method of claim 23 wherein the spectrum of the window function is approximately equal to zero at all integer non-zero multiples of the pitch frequency associated with the first frame.

25. (Original) The method of claim 2 wherein each pulse location corresponds to a time offset associated with an impulse in an impulse sequence, the first signal samples are computed by convolving the first digital filter with the impulse sequence, and the second signal samples are computed by convolving the second digital filter with the impulse sequence.

26. (Original) The method of claim 25 wherein the first signal samples and the second signal samples are combined by first multiplying each by a synthesis window function and then adding the two together.

27. (Original) The method of claim 1 wherein the spectral information includes a set of spectral magnitudes representing the speech spectrum at integer multiples of a fundamental frequency.

28. (Original) The method of claim 1 wherein the speech model parameters are generated by decoding a bit stream formed by a speech encoder.

29. (Original) The method of claim 1 wherein the first digital filter is computed as the product of a periodic signal and a pitch-dependent window signal, and the period of the periodic signal is determined from the pitch information for the first frame.

30. (Original) The method of claim 29 wherein the spectrum of the pitch dependent window function is approximately equal to zero at all non-zero integer multiples of the pitch frequency associated with the first frame.

31. (Original) The method of claim 1 wherein the first digital filter is computed by:

determining FFT coefficients from the decoded model parameters for the first frame in frequency regions where the voicing state equals the selected voicing state;

processing the FFT coefficients with an inverse FFT to compute first time-scaled signal samples;

interpolating and resampling the first time-scaled signal samples to produce first time-corrected signal samples; and

multiplying the first time-corrected signal samples by a window function to produce the first digital filter.

32. (Original) The method of claim 31 wherein regenerated phase information is computed using the decoded model parameters for the first frame, and the regenerated phase

information is used in determining the FFT coefficients for frequency regions where the voicing state equals the selected voicing state.

33. (Original) The method of claim 32 wherein the regenerated phase information is computed by applying a smoothing kernel to the logarithm of the spectral information for the first frame.

34. (Original) The method of claim 32 wherein further FFT coefficients are set to approximately zero in frequency regions where the voicing state does not equal the selected voicing state or in frequency regions outside the bandwidth represented by speech model parameters for the first frame.

35. (Original) The method of claim 31 wherein the window function depends on the decoded pitch information for the first frame.

36. (Original) The method of claim 35 wherein the spectrum of the window function is approximately equal to zero at all integer non-zero multiples of the pitch frequency associated with the first frame.

37. (Original) The method of claim 1 wherein the digital speech samples corresponding to the selected voicing state are further combined with other digital speech samples corresponding to other voicing states.

38. (Original) A method of decoding digital speech samples corresponding to a selected voicing state from a stream of bits, the method comprising:

dividing the stream of bits into a sequence of frames, wherein each frame contains one or more subframes;

decoding speech model parameters from the stream of bits for each subframe in a frame, the decoded speech model parameters including at least pitch information, voicing state information and spectral information;

computing a first impulse response from the decoded speech model parameters for a subframe and computing a second impulse response from the decoded speech model parameters for a previous subframe, wherein both the first impulse response and the second impulse response correspond to the selected voicing state;

computing a set of pulse locations for the subframe;

producing a set of first signal samples from the first impulse response and the pulse locations; and

producing a set of second signal samples from the second impulse response and the pulse locations; and

combining the first signal samples with the second signal samples to produce the digital speech samples for the subframe corresponding to the selected voicing state.

39. (Original) The method of claim 38 wherein the digital speech samples for the subframe corresponding to the selected voicing state are further combined with digital speech samples for the subframe representing other voicing states.

40. (Currently Amended) The method of claims claim 39 wherein the voicing state information includes one or more voicing decisions, with each voicing decision determining the voicing state of a frequency region in the subframe.

41. (Original) The method of claim 40 wherein each voicing decision determines whether a frequency region in the subframe is voiced or unvoiced.

42. (Currently Amended) The method of claims claim 41 wherein the pulse locations are reinitialized if consecutive frames or subframes are predominately not voiced, and future

determined pulse locations do not substantially depend on speech model parameters corresponding to frames or subframes prior to such reinitialization.

43. (Original) The method of claim 41 wherein each voicing decision further determines whether a frequency region in the subframe is pulsed.

44. (Original) The method of claim 41 wherein the selected voicing state is the voiced voicing state and the pulse locations depend at least in part on the decoded pitch information for the subframe.

45. (Currently Amended) The method of claims claim 44 wherein the pulse locations are reinitialized if consecutive frames or subframes are predominately not voiced, and future determined pulse locations do not substantially depend on speech model parameters corresponding to frames or subframes prior to such reinitialization.

46. (Original) The method of claim 45 wherein the frequency responses of the first impulse response and the second impulse response correspond to the decoded spectral information in voiced frequency regions and the frequency responses are approximately zero in other frequency regions.

47. (Original) The method of claim 46 wherein each of the pulse locations corresponds to a time offset associated with each impulse in an impulse sequence, and the first signal samples are computed by convolving the first impulse response with the impulse sequence and the second signal samples are computed by convolving the second impulse response with the impulse sequence.

48. (Original) The method of claim 47 wherein the first signal samples and the second signal samples are combined by first multiplying each by a synthesis window function and then adding the two together.

49. (Original) The method of claim 43 wherein the selected voicing state is the pulsed voicing state, and the frequency response of the first impulse response and the second impulse response corresponds to the spectral information in pulsed frequency regions and the frequency response is approximately zero in other frequency regions.

50. (Original) The method of claim 43 wherein the first impulse response is computed by:

determining FFT coefficients for frequency regions where the voicing state equals the selected voicing state from the decoded model parameters for the subframe;

processing the FFT coefficients with an inverse FFT to compute first time-scaled signal samples;

interpolating and resampling the first time-scaled signal samples to produce first time-corrected signal samples; and

multiplying the first time-corrected signal samples by a window function to produce the first impulse response.

51. (Original) The method of claim 50 wherein the interpolating and resampling the first time-scaled signal samples depends on the decoded pitch information of the first subframe.

52. (Currently Amended) The method of claims claim 51 wherein the pulse locations are reinitialized if consecutive frames or subframes are predominately not voiced, and future determined pulse locations do not substantially depend on speech model parameters corresponding to frames or subframes prior to such reinitialization.

53. (Original) The method of claim 51 wherein regenerated phase information is computed using the decoded model parameters for the subframe, and the regenerated phase information is used in determining the FFT coefficients for frequency regions where the voicing state equals the selected voicing state.

54. (Original) The method of claim 53 wherein the regenerated phase information is computed by applying a smoothing kernel to the logarithm of the spectral information.

55. (Original) The method of claim 53 wherein further FFT coefficients are set to approximately zero in frequency regions where the voicing state does not equal the selected voicing state.

56. (Original) The method of claim 55 wherein further FFT coefficients are set to approximately zero in frequency regions outside the bandwidth represented by decoded model parameters for the subframe.

57. (Original) The method of claim 51 wherein the window function depends on the decoded pitch information for the subframe.

58. (Original) The method of claim 57 wherein the spectrum of the window function is approximately equal to zero at all non-zero multiples of the decoded pitch frequency of the subframe.

59. (Currently Amended) The method of ~~claims~~ claim 38 and wherein the voicing state information includes one or more voicing decisions, with each voicing decision determining the voicing state of a frequency region in the subframe.

60. (Original) The method of claim 59 wherein each voicing decision determines whether a frequency region in the subframe is voiced or unvoiced.

61. (Currently Amended) The method of ~~claims~~ claim 60 wherein the pulse locations are reinitialized if consecutive frames or subframes are predominately not voiced, and future determined pulse locations do not substantially depend on speech model parameters corresponding to frames or subframes prior to such reinitialization.

62. (Original) The method of claim 60 wherein each voicing decision further determines whether a frequency region in the subframe is pulsed.

63. (Original) The method of claim 60 wherein the selected voicing state is the voiced voicing state and the pulse locations depend at least in part on the decoded pitch information for the subframe.

64. (Currently Amended) The method of ~~claims~~ claim 63 wherein the pulse locations are reinitialized if consecutive frames or subframes are predominately not voiced, and future determined pulse locations do not substantially depend on speech model parameters corresponding to frames or subframes prior to such reinitialization.

65. (Original) The method of claim 63 wherein the frequency responses of the first impulse response and the second impulse response correspond to the decoded spectral information in voiced frequency regions and the frequency responses are approximately zero in other frequency regions.

66. (Previously Presented) The method of claim 65 wherein each of the pulse locations corresponds to a time offset associated with each impulse in an impulse sequence, and the first signal samples are computed by convolving the first impulse response with the impulse sequence

and the second signal samples are computed by convolving the second impulse response with the impulse sequence.

67. (Original) The method of claim 66 wherein the first signal samples and the second signal samples are combined by first multiplying each by a synthesis window function and then adding the two together.

68. (Original) The method of claim 62 wherein the selected voicing state is the pulsed voicing state, and the frequency response of the first impulse response and the second impulse response corresponds to the spectral information in pulsed frequency regions and the frequency response is approximately zero in other frequency regions.

69. (Original) The method of claim 60 wherein the first impulse response is computed by:

determining FFT coefficients for frequency regions where the voicing state equals the selected voicing state from the decoded model parameters for the subframe;

processing the FFT coefficients with an inverse FFT to compute first time-scaled signal samples;

interpolating and resampling the first time-scaled signal samples to produce first time-corrected signal samples; and

multiplying the first time-corrected signal samples by a window function to produce the first impulse response.

70. (Original) The method of claim 69 wherein the interpolating and resampling the first time-scaled signal samples depends on the decoded pitch information of the first subframe.

71. (Currently Amended) The method of claims claim 70 wherein the pulse locations are reinitialized if consecutive frames or subframes are predominately not voiced, and future

determined pulse locations do not substantially depend on speech model parameters corresponding to frames or subframes prior to such reinitialization.

72. (Original) The method of claim 69 wherein regenerated phase information is computed using the decoded model parameters for the subframe, and the regenerated phase information is used in determining the FFT coefficients for frequency regions where the voicing state equals the selected voicing state.

73. (Original) The method of claim 72 wherein the regenerated phase information is computed by applying a smoothing kernel to the logarithm of the spectral information.

74. (Original) The method of claim 72 wherein further FFT coefficients are set to approximately zero in frequency regions where the voicing state does not equal the selected voicing state.

75. (Original) The method of claim 74 wherein further FFT coefficients are set to approximately zero in frequency regions outside the bandwidth represented by decoded model parameters for the subframe.

76. (Original) The method of claim 69 wherein the window function depends on the decoded pitch information for the subframe.

77. (Original) The method of claim 76 wherein the spectrum of the window function is approximately equal to zero at all non-zero multiples of the decoded pitch frequency of the subframe.